DOCUMENT RESUME

ED 048 595 AL 002 782

AUTHOR Pak, Tae-yong
TITLE The Universal Basis of the Sentence.
PUB DATE Apr 71
NOTE 19p.; Paper presented at the International Communication Association, Phoenix, Arizona, April 22-24, 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Componential Analysis, *Deep Structure, Korean, Language Universals, *Linguistic Theory, *Sentences, *Structural Analysis

ABSTRACT

This paper shows that the proposition, our universal thought process, underlies its linguistic realization, the sentence of a specific language, and provides the theoretical basis for interlingual translation as well as intralingual paraphrase. (An example of componential analysis is shown for some Korean sibling terms.) (Author/AMM)

ED 048595

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

THE UNIVERSAL BASIS OF THE SENTENCE

by

Tae-yong Pak

Department of English

University of Hawaii

Honolulu, Hawaii 96822

International Communication Association

Phoenix, Arizona

April 22 - 24, 1971

AL 002 782

5/3/71

THE UNIVERSAL BASIS OF THE SENTENCE

This paper shows that the proposition, our universal thought process, underlies its linguistic realization, the sentence of a specific language, and provides the theoretical basis for interlingual translation as well as intralingual paraphrase.

1. The proposition and the sentence

The proposition or complete thought has been the province of logic and philosophy, and the sentence that of linguistics. Unfortunately, few philosophers have shown sustained interest in showing how the sentence embodies the proposition, and few linguists, especially under the influence of formalistic structuralism, have sought to derive the sentence from the proposition, believing that their task consisted in mere enumeration of the vocabulary of a language and specification of their well-formed combinations, called the sentence. The cleavage thus created between meaning and form, between proposition and sentence, badly needs closing. There is of course nothing wrong with formalism itself and the linguist's preoccupation with the recurrence of words and their strings as physical shapes rather than as meaningful expressions parallels a similar impulse to be found in other disciplines upon reaching a certain stage of development (Wang 1955). What is objectionable is the sort of pseudo-formalism espoused by many linguists, for it is incapable of fully accounting for the facts of natural language.

Pseudo-formally viewed, a sentence is a string of words. Since the number of words is finite, it is conceivable that the denumerable set of sentences may somehow be enumerated. Such an enumeration can be facilitated by categorizing certain words as e.g. noun and verb according to their positions of occurrence within the sentential string. If all the rules which are necessary for the members of the various categories to form sentences are spelled out, it seems possible to present generative schemata (formation rules) for a natural language, say English, e.g. S = N + V + N.

The formational part of various axiomatic systems can be given without interpretation. There are various ways of making this fact intuitively clear. Perhaps the best-known method is to replace each primitive symbol by a GÖdel number and to consider the strings of symbols as products of the GÖdelized factors (GÖdel 1931). Such arithmetization certainly desemanticizes the formation part. Essentially the same principle is behind Quine's protosyntax (Quine 1962:283-318). The primitive elements are considered as typographical shapes, S_1, S_2, ... and their strings as concatenations of these shapes. Formalists assert that this sort of uninterpreted well-formedness should be the principal subject matter of linguistics, the study of natural languages.

It seems plausible to GÖdelize or devise a protosyntax for the elementary part of a natural language. We can regard the lexical items as meaningless objects, m_1, m_2, ... and consider all the statements concerning their well-formed combinations as proto-

syntactical, that is, statements concerning the physical strings of these objects. Even in an uninterpreted logistic it is customary to group the primitive elements into terms and predicates (Church 1951). However, this grouping does not presuppose any interpretation of the symbols and is intended merely to facilitate the statement of the combinatory rules. To appease the skeptic one may avoid in this categorization such semantic words as 'terms' and 'predicates' and use instead labels like Class One, Class Two, etc. in the manner of Fries (1952). For easy association, however, one may retain the traditional categories like noun and verb, provided that thereby no semantic interpretation is intended. Furthermore, the verbs should be subcategorized (in the sense of Chomsky 1965) according to the number of nouns of which the verbs predicate a certain property or relation. Thus some will be marked as intransitive, some as transitive, and so forth. Note here that no semantic interpretation may be intended. It is conceivable to group the verbs into V_1 and V_2, with no meaning. The subcategorization is performed only with a view to the forthcoming formation rules which may take the schematic form, 'N + V is well-formed, if V is from the list V_1; N + V + N is well-formed if the V is from the list V_2, etc.' Thus the expressions 'property', 'relation', 'predicate' are a mnemonic paraphrase of the protosyntactical predicate 'concatenation', expressed in the preceding sentence by the plus sign and the type of the primitive item marked by the set labels N and V.

But the formation rules following subcategorization, e.g. 'N + V is well-formed if V is from the list V_1', may generate grammatically correct but semantically deviant sentences, such as The boy lowed. Hence the need for 'selectional rules', formalized by Katz and Fodor (1963) and Chomsky (1965). These rules will reject e.g. *The boy lowed, although low is in the subcategory V_1, because low requires a bovine subject. Note that the selectional restrictions can be stated mechanically, up to a point. For example, a principle of uniform marking ($+F_i$, $-F_i$) may be adopted: each lexical item may be marked as to a certain feature, e.g. bovine, and only the strings with uniform + or - regarding that feature may be considered well-formed. The feature bovine is another proto-syntactical abbreviation, for the formation rules could have simply listed all the well-formed formulas in which low occurs in the V position: the ox lowed; the cow lowed; cattle lowed, etc. To characterize this set of well-formed formulas, the lexical items occurring in them may be arbitrarily marked, e.g. +bovine, without thereby intending any semantic interpretation.

But this encounters problems at once in conjoined nouns like my father and my mother. The only way to solve this is to admit that the conjuncts refer to different individuals, that there is no contradiction therefore, thus introducing semantics. But even this may seem avoidable by subscripting each conjunct in the manner of Langendoen (1969:48). However, the question of reference cannot be dismissed so lightly. In English, as in many other languages,

reflexivization or equi-NP deletion hinge on identical reference, and here certainly the mere inscriptional identity of forms is not enough, e.g. George (Smith) and George (Miller). The only way to accommodate this mechanically is to subscript each noun as many times as there are its referents. But pseudo-formalism has not carried out this necessary subscripting and sought to account for the communicative function of natural language by attending merely to inscriptional or phonetic identity.

The catastrophic consequences of pseudo-formalism have recently been shown in discourse analysis (Pak 1970). Harris (1963) claims that he can analyze a text of discourse by examining only the distributional data, 'without bringing into account other types of data, such as relations of meanings throughout the discourse'. But his method of linking recurrent segments in equivalence chains does not work because under his pseudo-formalistic orientation there is no way to determine the identity or equivalence of two phonetic configurations A and B.

How do we change pseudo-formalism into genuine, workable formalism? By carrying out the subscripting according to reference. That is, the word man in its various uses of nominal reference should not be treated as one and the same item from a crudely simplified lexicon but as thousands of distinct items, one for each human being, living and dead. But once we do this, we have none other than semantics and a sentence is a sentence, not because it fulfills certain concatenational rules for meaningless items in a lexicon but because each part refers to

something, i.e. has a meaning, and the whole expresses a complete thought, that is, a proposition. Even here one may conceivably maintain formalism, divorced from meaning, and consider subscripting a mere technicality. However, this brand of formalism is interchangeable with semantics and does not suffer from the kind of petrification pseudo-formalism imposes upon itself. But the question still remains: how are we going to explain the rather curious one-to-one correspondence between the formalistic terms and their referents, e.g. between $\underline{\text{man}}_{2,917,118,512}$ and George Miller?

Formalism, even when thorough, does not explain the motivation for the 'game' after formation and transformation rules. As a matter of fact, formalists implicitly assume the semantic motivation, despite protestations to the contrary. Plainly, words are not mere blocks or pieces but stand for some concepts, and sentences are not mere concatenations but stand for our thoughts. We must therefore ask for the meaning of the sentence, the extralinguistic entity called the proposition.

2. Synonymy and translation

There must be the proposition as a tertium quid when we speak of intralingual synonymy between two sentences or of interlingual translation. For the clarification of this proposition we must go outside language and consult our sense experience and conceptual thinking. Propositional activity goes on in every human thought and

the sentence in a specific language is articulation or linguistic embodiment of the proposition, which has the form

(1) $P^n (x_1, x_2, \ldots x_n)$

That is, we identify an n-tuple of particulars and assign the n-tuple to an n-place predicate. Thus in every proposition two complementary functions are at work, identification and predication (Strawson 1967: 1-17).

For the first function, identification, it may seem that the human perceiver is passive: he is merely subjected to sensory stimuli and all that is required of him is attention. However, this is not quite the case. We must perceive the stimuli at least as distinct and unique, or there will be chaos. Thus identification presupposes two basic notions, variability of stimuli and uniqueness of each stimulus. The first calls for variables and their universe and the second for quantification and, ultimately, class logic (Quine 1950:232). The question may arise: whence this sophisticated logical machinery which seems to precede even our first sensory exposure to the world? The natural answer seems to be that it is innate, but this study is concerned with the form, not the genesis, of thought and it suffices to state that wherever there is thought, logic is found as its very framework. Furthermore, we need not be metalogically aware of it, though we may embody it in every act of our thought, just as we may breathe and carry on the intricate metabolic functions without any knowledge of them.

For the other function, predication, the conceptual genius of man is actively at work, and perception of a predicate P applicable to a particular n-tuple, thus placing the n-tuple in the class P of similar n-tuples, is an original invention. When the concept P is based on ostensive exemplification, which is necessarily open-textured (cf. Waismann 1945), then it may be called empirical or 'ostensive' in Körner's terminology (1959:51-61). If for some reason it is convenient to simplify the concepts, they become closed (exhaustively and recursively definable) as in mathematics. The originality of our conceptualization accounts for the fact that given an n-tuple, we apply affirmatively to it predicates P_a, P_b, ... or negatively $P_{a'}$, $P_{b'}$, ... We may also have in view other concepts $P_{a''}$, $P_{b''}$, ... whose applicability to the n-tuple in question is a matter of contingency. Then the various predicates are related to each other either exactly as in Boolean algebra if they are simplifying or inexactly if empirical. The latter case comprehends, in addition to inclusion, exclusion and overlap, inclusion-or-overlap and exclusion-or-overlap (Körner 1959:41). The sum total of these first-order relations constitutes our primitive linguistic commitments in Körner's sense (1959:64) or 'meaning postulates' (Carnap 1956:222-229). The second or higher order logic based on these primitive commitments is indifferent to the actual contents of the commitments and merely gives us quantification and class logic (which is universal), obtaining among the committed relations of first order.

The primitive commitments (e.g. man is rational, 'red' is a color) in any language are numerous and complicated, but can be exhibited more or less economically as the existence of various lexicons attests.

The principle involved is called componential analysis (cf. Bendix 1966): the concepts to be distinguished are separated as bundles of binary semantic features. Theoretically, w concepts can be represented by $\log_2 w$ binary features, but just as considerations of naturalness increase the number of distinctive features in phonology above the minimum (Pak forthcoming (a) and (b)), the number of semantic features used by lexicographers is larger than $\log_2 w$, so that among the features we find inclusion (e.g. animateness includes human) or overlap (e.g. male crosscutting human and nonhuman). See Langendoen (1969:34-40).

In semantics, as in phonology, naturalness is largely arbitrary and idiosyncratic, and the isolation of certain features is conditioned by grammatical practice. In English we isolate gender and person, redundantly in case the predicates in the language are already distinguished by other binary features, because the two concepts have syntactic consequences, e.g. he vs she; he, she vs it; which vs who, etc. Similarly Berlin (1968) isolates some 250 Tzeltal counters, which undoubtedly overlap redundantly, for 2^{250} is many billions of times the actual size of Tzeltal vocabulary. An example of componential analysis is shown below for some Korean sibling terms using only the two features male and older: the first monadic and the second dyadic.

	Hyong	Tongsaeng	Nuna	Nui	Oppa	Onni
	$[x_1, x_2]$	$[x_1, x_2]$	$[x_1, x_2]$	$[x_1, x_2]$	$[x_1, x_2]$	$[x_1, x_2]$
F_1: male	[+ +]	[$\frac{+}{-}$ $\frac{\pm}{-}$]	[- +]	[- +]	[+ -]	[- -]
F_2: older	+[ ]	-[ ]	+[ ]	-[ ]	+[ ]	+[ ]

Table 1

where the dyad $[x_1, x_2]$ is related as 'x_1 is the P of x_2' or 'x_1 is F than x_2'; and F_2 applies to the dyad linearly, left to right, if + is prefixed to the brackets and conversely if otherwise. The kinship terms may be called nouns but are predicates like other verbs, except that they require a certain copula, e.g. be in English and i-(da) in Korean. These predicative nouns are to be distinguished from the identificatory (referentially used) nouns, which are derived by nominalization or relativization from the former. For details see Pak (forthcoming (c)).

If any two items, e.g. hyong and hyongnim, receive the same + or - markings for all relevant features, we find intralingual synonymy. However, this condition, 'all relevant features', should be properly understood. If they are to include prosodic or phonetic identity, then surely there can be no synonymy. In fact, the concept of synonymy excludes phonetic identity. But even after relaxation of this condition, differences in nuance, emotional connotation, etc. are bound to subsist and we must decide to disregard certain features if we are to have synonymy at all. Thus we pronounce synonymy at a certain level such as cognitive or emotive, circumscribing the sphere of our interest or 'relevant features'.

The above has been concerned with intralingual synonymy. Note that (a) the semantic features and (b) their bundling are language-specific. This again parallels the situation with phonological distinctive features, which to be sure make use of the universal vocal organs but which utilize language-unique loci, e.g. alveolar stop (Eng. /t/)

rather than postdental stop (Korean /t/), and their idiosyncratic uses, e.g. degree of tensity in occlusion and release. That (b) is idiosyncratic will become clear by comparing Table 1 with Table 2, which gives an analysis of the English predicates, brother and sister, in terms of the same semantic features.

	Brother	Sister
	$[x_1, x_2]$	$[x_1, x_2]$
F_1: male	[+ ±]	[- ±]
F_2: older	±[ ]	±[ ]

Table 2

However, (a), that is, the number and content of the semantic features, is no less language-specific. Surely such features as male and older will be found in all languages with no appreciable difference of content. However, the functional load of a feature may be different considerably. For instance, older is a frequently employed feature in languages with honorifics and perhaps this explains its irrelevance to the English kinship terms brother and sister. Furthermore, a feature like mythological, clear though it may appear, may not be operational in a culture in which distinction between mythology and history, superstition and science, is simply not relevant. The various ways in which color words divide the spectrum of visible light waves may be considered another example of (a) being distinct, arbitrarily.

How then do we expect to translate anything from one language into another. Two ways are open to the translator. If the features are functional in both languages and their semantic fields and weights are identical (as roughly between the features of Tables 1 and 2), while the predicates differ in their composition only in point of positive or negative presence of the features, one may superpose the missing $+F_i$ or $-F_i$ on to the predicate until equivalence is reached. For example, hyong in Korean may be translated into older brother of a male in English, and brother into something like hyong-ina-tongsaeng-ina-oppa where ina is roughly equivalent to or. If the features are not functional in the language, then the concept must be identified by locutions like P such that ...P..., where ...P... is a sentence using the predicate P. The P, being nameless,

must be a proverb form like Eng. be or do, provided that the ordered set of particulars to which the predicate applies is identifiable. And this is possible in any language so long as there are the demonstratives. That is why I claim to show complete and determinate translatability (Pak 1971), notwithstanding Quine's theory of indeterminacy (1960:26-79).

3. Conclusion

In the canonical form (1) of the proposition, P^n, that is, P_1^1, P_2^1, ... P_1^2, P_2^2, ... P_1^n, P_2^n, ..., constitute the lexicon of predicates, which pseudo-formalism mistakes to be the complete lexicon. On the other hand, the x's or particulars are the stimuli and objects that impinge on our senses, which are identified by various linguistic devices using the predicates. The details whereby (1) becomes realized as the actual English sentence, of the schema

$$(2)\ \text{Subj} + \text{Aux} + \text{Pred}_1 + \left[\begin{array}{l} \left[\begin{array}{l} (\text{Obj}_1) + \left[\begin{array}{l} (\text{Obj}_2) \\ (\text{Pred}_2) \\ (\text{Adv}) \end{array}\right] \\ (\text{Pred}_2) \end{array}\right] \\ (\text{Adv}) \end{array}\right] + (\text{Adv})$$

are beyond the scope of this paper and the reader is referred to Pak (forthcoming (c)), where I show that not only Subj and Obj but Aux and Adv are particulars, that their linguistic realization is achieved by processes known as nominalization and relativization, corresponding, respectively, to 'description' (Quine 1950:216) and, partly, 'abstraction' (Quine 1950:242) in logic. It goes without saying that the particular, nominalized or relativized, is an identified individual in reality. The linguistic expressions Subj, Obj, etc. refer to the particulars and are not mere typographical forms.

REFERENCES

Bendix, Edward Herman. 1966. Componential analysis of general vocabulary. IJAL 32.

Berlin, Brent. 1968. Tzeltal numeral classifiers: a study in ethnographic semantics. Janua linguarum, series practica 70. Hague: Mouton.

Carnap, Rudolf. 1956. Meaning and necessity. Chicago and London: Chicago U. P.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, Mass.: MIT Press.

Church, Alonzo. 1951. The need for abstract entities in semantic analysis. American Academy of Arts and Sciences Proceedings, Vol. 80. 100-112.

Fries, Charles C. 1952. The structure of English. New York: Harcourt, Brace and World.

Gödel, Kurt. 1931. Uber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte für Mathematik und Physik. Vol. 38. 173-198.

Harris, Zellig S. 1963. Discourse analysis reprints. Hague: Mouton.

Katz, Jerrold J. and Jerry A. Fodor. 1963. The structure of a semantic theory. Language 39. 170-210.

Körner, Stephan. 1959. Conceptual thinking: a logical inquiry. New York: Dover.

Langendoen, D. Terence. 1969. The study of syntax. New York, etc.: Holt, Rinehart and Winston.

Pak, Tae-yong. 1970. Review of Discourse analysis reprints by Zellig S. Harris. Language 46. 754-764.

Pak, Tae-yong. Forthcoming (a). Convertibility between distinctive features and phonemes. Linguistics.

_____. Forthcoming (b). The economy of distinctive features in phonology. Linguistics.

_____. Forthcoming (c). A logical grammar of Modern English. Working Papers in Linguistics. University of Hawaii.

_____. 1971. Complete and determinate translatability. Read at the Pacific Conference on Contrastive Linguistics and Language Universals, January 1971, University of Hawaii.

Quine, W. V. 1950. Methods of logic. New York: Holt.

_____. 1960. Word and object. Cambridge, Mass.: MIT Press.

_____. 1962. Mathematical logic. Revised ed. New York: Harper and Row.

Strawson, P. F. 1967. Introduction. Philosophical logic, ed. by P. F. Strawson, 1-16. London: Oxford U. P.

Waismann, Friedrich. 1945. Verifiability. Proceedings of the Aristotelian Society, Supp. Vol. 19. 119-50. Reprinted in The theory of meaning, ed. by G. H. R. Parkinson, 1968. 35-60. London: &xxx&xEE Oxford U. P.

Wang, Hao. 1955. On formalization. Mind 64. 226-238.